

Variações de Mensuração por Tipos de Escalas de Verificação: Uma Análise do Construto de Satisfação Discente

Variations of Measurement by Types of Verification Scales: An Analysis of the Student Satisfaction Construct

Francisco José da Costa¹, Anna Carolina Rodrigues Orsini¹, Jailson Santana Carneiro¹

¹Universidade Federal da Paraíba, UFPB, Brasil.

Correspondência: Anna Carolina Rodrigues Orsini. Cidade Universitária, CEP 58.051-900, Paraíba, PB, Brasil.
Telefone: +55 (83) 3216-7200. E-mail: carolorsini@hotmail.com.

Recebido: 17 de setembro de 2017 Aceito: 17 de outubro de 2018 Publicado: 28 de dezembro de 2018

DOI: <http://dx.doi.org/10.21714/1679-18272018v16n2.p132-144>

Resumo

A temática de mensuração no campo de ciências sociais é recorrente em termos de indicação da necessidade de contribuições que minimizem erros relacionados ao uso de métricas em pesquisas quantitativas. O objetivo desse estudo foi analisar, por medidas descritivas e testes estatísticos de comparação, os efeitos do uso de oito diferentes escalas de verificação, com formatos binários, de quatro, cinco e sete pontos. A análise foi desenvolvida com base no construto de satisfação discente, para qual foi feita uma survey com alunos de duas instituições públicas de ensino superior. Em geral, os resultados permitiram verificar nuances, variações e vantagens ou desvantagens dos diferentes tipos de escalas, considerando objetivos dos estudos, temáticas e posições do pesquisador. O estudo traz contribuições na esfera metodológica das pesquisas relacionadas ao campo de ciências sociais e comportamentais, bem como em esfera prática profissional e acadêmica.

Palavras-chave: Escalas, Mensuração, Satisfação.

Abstract

The measurement theme in the field of social sciences is recurrent in terms of indicating the need for contributions that minimize errors related to the use of metrics in quantitative research. The objective of this study was to analyze through descriptive measures and comparative statistical tests, the effects of the use of eight different check scales, with binary formats, of four, five and seven points. The analysis was developed based on the student satisfaction construct, for which a survey was conducted with students from two public institutions. In general, the results allowed to verifying nuances, variations and advantages or disadvantages of the different types of scales, depending on the objectives, themes and positions of the researcher. The study brings contributions in the methodological sphere of research related to the field of social and behavioral sciences, as well as in professional and academic sphere of practice.

Keywords: Scales, Measurement, Satisfaction.

Esta obra está licenciada sob uma Licença Creative Commons Attribution 3.0.

1. Introdução

Esse estudo analisa usos de escalas de verificação, buscando apresentar uma contribuição para o campo de teoria da mensuração no contexto brasileiro. A teoria de mensuração envolve o esforço de compreensão sobre algum objeto e produção de conhecimento sobre um construto definido e suas relações com outros construtos (MARI, 2003). Desse modo, a produção de conhecimento para fins acadêmicos ou profissionais passa pelo aperfeiçoamento contínuo dos esforços de medição, principalmente em pesquisas quantitativas, que se baseiam em técnicas estatísticas aplicadas a números extraídos de objetos e associados a construtos (SILVA JÚNIOR; COSTA, 2014).

O tema 'mensuração' está diretamente relacionado com o estabelecimento de escalas de mensuração, o que abarca decisões de métrica (escala de verificação ou atribuição), de regras de implementação e de referenciais gerais. No

que se refere ao uso de escalas, existem diversas formatações desenvolvidas com o intuito de aprimorar a captação de aspectos latentes importantes para o desenvolvimento do conhecimento científico e prático em várias áreas acadêmicas e profissionais (cf. COSTA, 2011; LUCIAN, 2016).

Alguns desdobramentos e pesquisas buscaram contribuir para o desenvolvimento do campo reconhecendo limitações e vantagens quanto ao uso de determinados tipos de escala (DOLNICAR; GRÜN, 2013), sugerindo procedimentos de elaboração discutindo comparativamente seus desempenhos (cf. BARBOZA et al., 2013; CARVALHO; COSTA; SOUZA, 2015) e refletindo sobre questões específicas como, por exemplo, a influência de pontos neutros nas escalas ou variações de pontos (cf. GARLAND, 1991). Em geral, esses estudos são orientados à definição de procedimentos melhores fundamentados para uso de escalas mais válidas e confiáveis.

Mesmo sendo tema recorrente, várias questões ainda são postas quando buscamos alcançar medidas consistentes e adequadas aos diferentes objetos de estudo e respectivos contextos de pesquisa. Nesse sentido, o objetivo desse estudo foi analisar os efeitos do uso de diferentes escalas de verificação em pesquisas quantitativas, com base em um levantamento empírico.

Para alcançar esse objetivo, inicialmente apresentamos uma sessão de apontamentos teóricos em que tratamos sobre o problema da mensuração, os tipos de escalas de verificação e suas variações e implicações de uso para as pesquisas quantitativas. Em seguida, detalhamos os procedimentos metodológicos adotados para construção do estudo. No tópico subsequente, destacamos os resultados alcançados a partir da análise das escalas testadas e, por fim, apresentamos breves considerações finais, destacando os principais impactos, limitações e possíveis desdobramentos futuros.

2. Apontamentos Teóricos

Neste item estão dispostos os elementos teóricos sobre a temática base do artigo. Apresentamos, inicialmente, questões relativas ao problema de mensuração, especialmente nas ciências sociais e comportamentais. Em seguida, discutimos os tipos de escalas de verificação e, por fim, os formatos possíveis de escalas de verificação para estudos quantitativos.

Cabe adiantar um esclarecimento sobre os usos da palavra escala, e utilizaremos aqui a forma proposta por Costa (2011), para quem a palavra escala pode, inicialmente, ter o sentido de instrumento de medição (envolvendo os itens de verificação, os números de verificação de intensidade ou classificação e, por fim, o conjunto de regras de aplicação do instrumento), e ainda se associar aos números da escala (por exemplo, as escalas de concordância de 1 a 5, de 1 a 7, de 1 a 10). Para distinção de uso, o autor sugere que esse segundo sentido seja identificado por 'escala de verificação', em contraste com a 'escala de mensuração', que envolve o conjunto completo de elementos.

2.1. O Problema Da Mensuração

A mensuração consiste na representação de atributos de objetos e eventos do mundo real através de símbolos (FINKELSTEIN, 2009). A problemática da mensuração se intensifica a partir de sua aplicação a construtos abstratos (prazer, satisfação, por exemplo), que são comuns nas ciências sociais e comportamentais. Um pressuposto fundamental da teoria da mensuração é o de que qualquer construto é mensurável, de modo que a natureza abstrata de determinados objetos ou construtos não é impedimento para sua mensuração (BARBOZA et al., 2013); o desafio, portanto, é de identificar uma forma viável de conceber e aplicar a melhor métrica.

O debate em torno de mensuração e construção de escalas já está bem avançado e os esforços se concentram em refletir sobre essa melhor métrica para mensurar a grande variedade de construtos relevantes. Especialmente no caso de construtos ditos latentes, que não podem ser acessados diretamente, Costa (2011) destaca que uma alternativa consiste em medir características ou manifestações observáveis do mesmo. Por exemplo, podemos medir o nível de 'lealdade' de um consumidor a uma marca verificando a quantidade de postagens positivas que ele faz sobre a marca em redes sociais.

As escalas de mensuração tornam-se ferramentas importantes para acessar essas manifestações ou características observáveis de construtos abstratos, captando suas manifestações da maneira mais precisa e com segurança de que a manifestação ou característica de fato representa o construto em análise. Assim, é possível transformar manifestações em números passíveis de análise e a posterior implementação de técnicas quantitativas de análise de intensidade, associações, relações de predição etc. dos construtos.

A utilização de escalas de aferição indireta, como no exemplo dado, tem se mostrado útil nas áreas de saúde (e.g. FEITOSA et al., 2014), de educação e de administração (cf. COSTA, 2011), e por essa razão, são comuns estudos refletindo sobre o assunto. Pesquisas recentes em torno desse tema têm explorado os tipos de métricas disponíveis para utilização em pesquisas e possíveis variações nos resultados decorrentes das escalas adotadas (DAWES, 2008; BARBOZA et al., 2013), além de questões relacionadas, como a utilização de ambiente virtual e presencial e a

aplicação de questionários impressos e online, inclusive em redes sociais virtuais (CARVALHO; COSTA; SOUZA, 2015; RIFE et al., 2014; WARD et al., 2012). O item seguinte discute sobre os tipos de escalas.

2.2. Tipos De Escalas De Verificação

Nesse item nosso foco são as modalidades de escalas de verificação. Desde o início do século XX, vários formatos específicos de escalas de verificação já foram desenvolvidos, sendo destacados, para a finalidade desse estudo, os formatos (e suas variações) da escala binária, da escala de concordância de Likert, da escala Phrase Completion e da escala de diferencial semântico (COSTA, 2011).

A escala binária é aquela em que o respondente é apresentado a um item e informa qual sua posição entre duas opções. Por exemplo, um consumidor é questionado se ficou ou não satisfeito com um serviço, sendo dada a opção de indicar 'sim' ou 'não'. Já a escala Likert é a de maior aplicação nas ciências sociais e consiste na atribuição de números associados a níveis de concordância com determinada afirmação relativa a um construto. As variações nas escalas Likert são diversas, mas usualmente as mais comuns são de 5 (de 1 a 5), 7 (de 1 a 7), 10 (de 1 a 10) e 11 (de 0 a 10) pontos (BARBOZA et al., 2013).

Um formato mais recente foi a escala phrase completion. Esta escala foi desenvolvida por Hodge e Gillespie (2003) como uma alternativa à escala Likert e suas dificuldades de aplicação e verificação, como por exemplo, a definição dos pontos em sentido de concordância nas escalas. Nessa escala, o respondente é apresentado a uma frase que ele 'completa' com a intensidade associada ao construto de referência. Esse referencial da escala é montado com onze pontos (0-10) (SILVA JÚNIOR; COSTA, 2014). Por fim, é recorrente o uso da escala de Diferencial Semântico, que foi proposta em 1957 por Osgood, Suci e Tannenbaum, e que consiste em uma escala cujos extremos são adjetivos antônimos, como, por exemplo, Péssimo-Excelente ou Fraco-Forte (ANDRADE et al., 2009). Esta é uma escala usada para verificar o significado atribuído pelos participantes a algo (frase, foto etc.). As sinalizações são de que o uso das escalas gera efeitos em termos de vantagens e desvantagens para os estudos na área.

2.3 Análise Do Efeito Da Escala Na Medição

Conforme mencionado, boa parte dos estudos sobre mensuração tem se concentrado na verificação de possíveis efeitos de variação entre as diversas escalas, dado que não há um consenso entre os estudiosos de qual escala tem melhor desempenho. Com efeito, assim como a escolha da teoria, também o tipo de escala em uma pesquisa pode afetar as informações extraídas de maneira significativa (LOPES; GIRALDI, 2014). Neste sentido, alguns estudos em nível de Brasil (e.g. BARBOZA et al., 2013; DALMORO; VIEIRA, 2013; SILVA JÚNIOR; COSTA, 2014) e internacionalmente (e.g. DAWES, 2008; UNNITHAN, 2016) já empreenderam esforços no intuito de verificar possíveis divergências entre as variações dos formatos.

O estudo de Barboza et al. (2013), por exemplo, verificou que o uso de escalas de Likert de 5, 7 e 11 pontos pode gerar bons resultados, independente do formato, apesar de existir diferenças estatísticas (pequenas) nas medidas descritivas. Por sua vez, Dalmoro e Vieira (2013) testaram escalas de 3 (1 a 3), 5 (1 a 5) e 7 (1 a 7) pontos e concluíram que a escala de cinco pontos foi mais adequada ao contexto de pesquisa.

Diante da variedade de opções e considerando as diferentes aplicações, Dolnicar e Grün (2013) realizaram um estudo mais abrangente, verificando possíveis influências de vários formatos de escalas com o intuito de analisar comportamentos e, assim, permitir uma melhor comparação entre estudos. Entre os resultados obtidos, há sinalização de que as respostas 'sim' e 'não' fornecidas em escalas binárias completas são mais 'traduzidas' em pontos menos extremos em escalas do tipo Likert, como por exemplo 'concordo' ou 'discordo', e não em 'concordo totalmente' e 'discordo totalmente'.

O estudo de Dolnicar e Grün (2013) não se limitou à escala Likert, e explorou uma gama maior de possibilidades como escala binária, além de variação nos números (-3 a +3, por exemplo). Desse modo, o estudo desses autores foi utilizado como base para a construção empírica da nossa pesquisa, de modo que também analisamos diferentes formatos de escalas para fins comparativos. No tópico seguinte descrevermos especificamente os procedimentos metodológicos adotados.

3. Procedimentos Metodológicos

Nesse item apresentamos os detalhes sobre as decisões e procedimentos do trabalho empírico realizado. Dada a finalidade de explorar variações por diferentes tipos de escalas de verificação, optamos então por considerar formas distintas de captação de posições dos respondentes, o que requereu a definição de respondentes, construtos e escalas.

Em termos de respondentes, decidimos que seriam estudantes universitários, dada a facilidade de acesso e o não

comprometimento que esse grupo de respondente gera para a pesquisa. Apesar de reconhecermos os potenciais vieses que respostas de estudantes geram em pesquisas empíricas (BURNET; DUNE, 1986; PETERSON, 2001), esse viés não parece ser problemático nessa pesquisa, cujo resultado está concentrado no comportamento de resposta das escalas e não em associações de construtos teóricos.

Com relação aos construtos a serem mensurados, a nossa decisão foi a de mensurar satisfação do discente e, para termos uma maior disponibilidade de referências de avaliação, optamos por analisar a satisfação não apenas em um, mas em três aspectos principais do serviço que recebem (ALCÂNTARA et al., 2012; CASTRO, 2012). Avaliamos então a satisfação quanto à estrutura (ambiente geral e infraestrutura do curso), quanto ao currículo (estrutura curricular do curso para formação) e quanto ao corpo docente. Essa decisão está de acordo com o que em geral se explora na literatura sobre esse construto multidimensional, recorrente em estudos nas áreas de administração e afins.

Por fim, em relação às escalas, optamos por utilizar 8 tipos das inúmeras possíveis, com a tentativa de selecionar aquelas opções que estão dentre as mais usadas em pesquisas acadêmicas, profissionais e de mercado no Brasil. O Quadro 1 indica os tipos de escalas e a forma de verificação.

Quadro 1: Escalas usadas no estudo empírico.

Escala	Explicação
1 - Binária por afirmação	Era solicitado que o respondente apenas marcasse, nas três dimensões avaliadas, o(s) aspecto(s) que estava satisfeito. A pergunta teve o seguinte enunciado: “Considerando as características abaixo relativas ao seu curso, aponte COM QUAIS você está SATISFEITO(A)”.
2 - Binária completa	Era solicitado que o respondente marcasse, em cada um dos três aspectos avaliados, com qual aspecto estava satisfeito apontando ‘sim’ ou ‘não’. A pergunta teve o seguinte enunciado: “Considerando as características abaixo relativas ao seu curso, aponte SE VOCÊ ESTÁ OU NÃO SATISFEITO(A)”
3 - Unipolar de 4 pontos	Era dado o seguinte enunciado: “Considerando as características abaixo de seu curso, aponte o quanto você está satisfeito(a), segundo as opções dadas”, e abaixo indicação do aspecto avaliado (Por exemplo, “Com a estrutura do curso”), e o respondente marcaria sua satisfação em 4 pontos: nada satisfeito(a); um pouco satisfeito(a); bastante satisfeito(a); extremamente satisfeito(a).
4 - Likert de 4 pontos verbal	Era feita uma afirmação de que o respondente estava satisfeito (por exemplo: “Sinto-me satisfeito com a estrutura do curso”), e eram dadas quatro níveis de concordância: discordo totalmente; discordo; concordo; concordo totalmente.
5 - Likert de 5 pontos não verbal	Era feita a uma afirmação (exemplo: “Sinto-me satisfeito com a estrutura do curso”) e o respondente marcaria um nível entre 5 opções (-2, -1, 0, +1, +2). O enunciado informava que os extremos indicavam discordância total (-2) e concordância total (+2)
6 - Likert de 5 pontos verbal	Era feita uma afirmação de que o respondente estava satisfeito (por exemplo: “Sinto-me satisfeito com a estrutura do curso”), e eram dados cinco níveis de concordância: discordo totalmente; discordo; nem concordo nem discordo; concordo; concordo totalmente.
7 - Bipolar de 7 pontos não verbal	Era dada uma frase relativa à dimensão de avaliação (por exemplo: “Em relação ao meu curso, considerando sua estrutura, estou...”), e dois extremos de resposta indicados como ‘Insatisfeito’ e ‘Satisfeito’. Entre esses extremos haviam 7 pontos, nessa sequência: -3, -2, -1, 0, +1, +2, +3.
8 - Bipolar de 7 pontos verbal	Era dada uma frase relativa à dimensão de avaliação (por exemplo: “Em relação ao meu curso, considerando sua estrutura, estou...”), e dois extremos de resposta indicados como ‘Insatisfeito’ e ‘Satisfeito’. Entre esses extremos haviam 7 âncoras, nessa sequência: Extremamente; Bastante; Um pouco; Nem um nem outro; Um pouco; Bastante; Extremamente

Fonte: Elaborado pelos autores (2017).

No desenho do trabalho de campo, os itens das escalas foram organizados em um instrumento de pesquisa impresso, que incluía diversas outras questões (sociodemográficas, escalas de outros construtos). Foram então construídos 8 tipos de instrumentos, sempre com duas páginas no documento impresso, e cada um continha duas dessas escalas, uma sempre no início e uma sempre no final. Por exemplo, um tipo de questionário trazia no começo a escala ‘Binária por afirmação’ e no final a ‘Bipolar de 7 pontos não verbal’; outro se diferenciava desse somente pela inversão dessa ordem, ou seja, no início a ‘Bipolar de 7 pontos não verbal’, e no final a ‘Binária por afirmação’. Considerando as numerações da primeira coluna do Quadro 1, os demais traziam ordenadamente, as seguintes escalas: 2 e 6; 6 e 2; 4 e 5; 5 e 4; 3 e 8; e 8 e 3. Essa estratégia foi adotada para termos a medição de dois tipos de escalas por questionário, sem, no entanto, trazer vieses relacionados à posição da escala no instrumento ou indispor o respondente à resposta pela percepção de perguntas equivalentes e repetitivas.

Na impressão dos instrumentos, os diferentes tipos de questionário foram então colocados em blocos de 8

instrumentos, segundo as sequências descritas no parágrafo anterior, de forma que, quando da aplicação, a distribuição no espaço de coleta seria de tal modo que todas as formas de escalas estariam sendo aplicadas e com dispersão tal que não se evidenciasse tendência de resposta que fosse decorrente, por exemplo, das características de alguma sala de aula, do turno de coleta, da instituição etc. Em termos de quantidade de respostas, nossa meta inicial foi ter ao menos 40 respondentes para cada tipo de instrumento, o que nos fez prever, portanto, uma amostra de 320 respondentes. Esse número de respondentes por tipo de instrumento foi definido de acordo com a necessidade de verificação consistente de resultados estatísticos das diferenças potenciais entre escalas.

Com esse desenho preliminar, empreendemos a coleta junto a discentes de cursos de Administração de duas instituições federais de ensino de um estado do Nordeste brasileiro. Decidimos por delimitar a coleta apenas ao curso de Administração com vistas a manter a homogeneidade quanto à compreensão do construto (satisfação) nas dimensões expostas (corpo docente, currículo e estrutura) e por se considerar um curso diversificado em termos de perfil discente, comparado a outros cursos. A coleta ocorreu no período de março a maio de 2017, e a amostra total foi de 329 respondentes. A amostra ficou mais concentrada do segundo ano de curso em diante (14,4% do primeiro ano, 24,8% do segundo ano, 36,4% do terceiro ano, e 24,4% dos demais anos), mais concentrados nos horários diurnos (56,5%). Em termos de idade, tivemos maior concentração na faixa de até 24 anos (67,2% do total) e, quanto ao estado civil, 79,6% declararam-se solteiros. Tivemos ainda um maior volume relativo de homens (54,4%) e uma concentração maior de pessoas com renda familiar até R\$3.000 reais mensais (57,8%). Os resultados sinalizam a obtenção de uma amostra com as características próximas do que se conhece (por experiência dos autores) de alunos em cursos de graduação em Administração de instituições públicas de ensino.

Em termos de distribuição dos instrumentos, as quantidades giraram entre 40 e 42 respondentes para cada tipo de questionário, muito próximo do que fora previsto. Como cada escala estava em dois dos instrumentos (um no começo e outro no final), tivemos então, por cada tipo de escala, totais variando entre 80 e 84 respondentes. Embora seja uma quantidade pequena, entendemos que o volume de respostas por escala permite a realização de conclusões consistentes sobre o desempenho e as diferenças de escalas.

Concernente à análise, optamos por proceder a algumas avaliações com técnicas descritivas, analisando as regularidades de comportamento por escala, todas realizadas por meio do software R. Dado o desenho apresentado, nossa expectativa é de que não há motivo para variação de nível de satisfação dos alunos por instrumento coletado, e, por isso, testaremos a independência das características da amostra em relação às diferentes variáveis categóricas utilizadas nos 8 tipos de instrumento, de modo a assegurar que qualquer variação identificada seja nas métricas, caso ocorra, será mais oriunda da composição das escalas.

4 Resultados

Conforme indicado, tivemos 8 tipos de questionários e o desenho de campo foi programado de tal forma que, numa mesma sala de aula, os instrumentos eram entregues em sequência para tornar aproximadamente homogêneos os tipos de respondentes para cada tipo e configuração de questionário. Assim, como primeira atividade de análise, optamos por testar pelo teste de qui-quadrado, aplicado em tabelas cruzadas, a independência do tipo de questionário em relação às variáveis categóricas de semestre, de curso, faixa de idade, faixa de renda, gênero e turno. A suposição é de que, caso se verifique independência do tipo de questionário em relação a essas variáveis, tem-se evidência de que a estratégia de distribuição não gerou viés decorrente das características da amostra.

Pelos testes, observamos que não há indício de dependência do tipo de questionário com qualquer das variáveis categóricas usadas (semestre - $\chi^2=58,1$, 77 gl [graus de liberdade], p-valor=0,947; idade - $\chi^2=21,14$, 21 gl, p-valor=0,450; renda - $\chi^2=25,36$, 21 gl, p-valor=0,232; gênero - $\chi^2=5,08$, 7 gl, p-valor=0,650; turno - $\chi^2=1,87$, 7 gl, p-valor=0,967). Ou seja, por esses resultados é possível afirmar que conseguimos isolar com sucesso o efeito da escala de outros possíveis condicionamentos das características das amostras selecionadas. Dessa forma, as variações que analisaremos, caso existam, serão atribuídas especificamente às características da escala.

4.1 Análise das respostas para escalas binárias

De acordo com o apontado no item 3, tivemos coleta por duas escalas binárias, uma na qual o respondente apenas marcava o item (estrutura, currículo e corpo docente) se estivesse satisfeito (binária por afirmação), e a outra em que ao lado de cada item eram apontadas as opções de ‘sim’ ou ‘não’ (binária completa). A Tabela 1 e o Apêndice A mostram os resultados, pelos quais se observa que, nas três variáveis, os percentuais das respostas positivas são sempre maiores na escala binária completa.

Tabela 1: Medidas por escalas binárias.

Opções	Binária completa		Binária por afirmação	
	Freq.	Percen.	Freq.	Percen.
Estrutura ($\chi^2 = 7,85, 1 \text{ gl}, p\text{-valor}=0,0051$)				
Sim	50	59,5	30	36,6
Não	34	40,5	52	63,4
(NM)*				
Total	84	100,0	82	100
Currículo ($\chi^2 = 1,17, 1 \text{ gl}, p\text{-valor} = 0,2787$)				
Sim	47	56,0	38	46,3
Não	37	44,0	44	53,7
(NM)				
Total	84	100,0	82	100,0
Corpo docente ($\chi^2 = 6,53, 1 \text{ gl}, p\text{-valor} = 0,0106$)				
Sim	63	75,0	45	54,9
Não	21	25,0	37	45,1
(NM)				
Total	84	100,0	82	100,0

Fonte: Dados da pesquisa (2017).

*NM =Não Marcou.

Considerando que temos duas amostras independentes de respondentes, optamos então por proceder ao teste das proporções para analisar a hipótese (nula) de que não há diferença estatística entre os percentuais de respostas positivas nos dois tipos de escalas. Após implementar o teste de proporções no pacote R, extraímos a estatística do teste e o p-valor, que estão indicados ao lado da indicação da variável, ainda na Tabela 1. A verificação que temos, ao um nível de significância de 5%, é de que em duas das três variáveis podemos efetivamente afirmar que a escala completa gera resultados de satisfação mais elevados (especificamente na variável ‘currículo’ é que não foi verificada diferença estatística, ainda que o nível de satisfação, tomando apenas a frequência, seja maior na escala binária completa).

Como primeira conclusão, podemos então afirmar que o nível declarado de satisfação (se considerarmos o percentual de satisfeitos como um indicador desse nível) é obviamente condicionado pelo formato da escala de verificação, de tal modo que a escala binária completa parece evidenciar níveis mais elevados de satisfação que a escala por afirmação. A diferença de resposta ocorre porque, possivelmente o respondente reflete mais quando tem que se manifestar positiva ou negativamente sobre sua satisfação. Assim, se não for requerido que aponte posição, ele pode simplesmente se negar a responder sem refletir se está ou não de fato satisfeito. Obviamente, a não resposta não significa insatisfação e, por essa razão, temos uma segunda conclusão de que, no levantamento de satisfação, a aferição parece ser melhor se o respondente for explicitamente solicitado a dar uma posição, sem a opção de ‘não resposta’.

4.2 Análise Das Respostas Para Escalas De Likert

Como indicado no método, usamos a escala de Likert em três formatos (4 e 5 pontos verbal, e 5 pontos não verbal). Os resultados para os percentuais e os gráficos estão no Apêndice B. Avaliando as frequências, chama a atenção o fato de não parecer haver proximidade de valores entre os pontos correspondentes nas escalas de cinco pontos. Na escala não verbal, por exemplo, o nível máximo de concordância (que era +2) recebe percentuais entre 15 e 20%, mas na escala verbal, o nível correspondente tem valores oscilando entre 2 e 11%, ou seja, sempre valores menores. Nessa segunda escala (a verbal), por outro lado, o ponto intermediário recebe percentuais mais elevados (em duas variáveis) que o correspondente na escala não verbal. A indicação reforça a conclusão já verificada em outros estudos (ver COSTA, 2011) de que em escala verbal com ponto neutro explícito, este ponto tende a absorver maior percentual de respostas, provavelmente por ser o ponto que o respondente não tem que se esforçar para analisar e responder.

A escala de 4 pontos, por outro lado, eliminou o ponto neutro. Pelo seu resultado, observamos que o ponto neutro parece ‘puxar’ respostas de discordância, uma vez que as marcações somadas de concordância se aproximam tanto das marcações explícitas de concordância da escala verbal quanto das marcações positivas da escala não verbal (os percentuais somados de concordância, respectivamente, para a escalas de cinco pontos verbal, não verbal, e de 4 pontos, foram: 59,1%, 54,2% e 55,4% para estrutura; 55,4%, 49,4% e 57,9% para currículo; e 56,7%, 57,3% e 61,5% para corpo docente). Quando comparamos esses resultados com os percentuais da escala binária completa da Tabela 1, que supomos ser a mais adequada para captação de satisfação, observamos que, com exceção da

variável ‘corpo docente’, os valores somados aproximam-se daqueles da resposta ‘sim’ da escala binária. Podemos então concluir que, em escalas de Likert, os valores relativos a concordância (que são os dois valores acima do ponto intermediário ou a metade superior das escalas de 4 pontos) correspondem, quando somados, ao percentual de indicação positiva de satisfação. Esse resultado é especialmente relevante em conversão de escala de intervalo para escalas binárias, o que é o caso, por exemplo, de adoção de técnicas de teoria da resposta ao item para escalas intervalares que são convertidas em binárias (cf. Alexandre et al., 2001).

Para uma análise comparativa complementar entre os três tipos de escalas, optamos por analisar as medidas descritivas de média e variância, o que nos levou a converter a escala de 4 pontos para o padrão de 1 a 5 (a fórmula de conversão foi: $N = 1 + 4(O - 1)/3$, em que N é o novo valor e O é o valor original na escala de 1 a 4). Os resultados de média, desvio padrão e quartis estão indicados na Tabela 2, pelos quais se observa que as escalas tendem a gerar valores de medidas de posição (média e quartis) e de dispersão muito próximos.

Tabela 2: Medidas por escalas de Likert.

Escalas	Média	Desvio	Quartis		
			Q1	Q2	Q3
Estrutura (Anova – F=0,927, 2 e 246 gl, p=0,397; Levene – F=1,251, 2 e 246 gl, p=0,288)					
Quatro pontos verbal	3,15	1,01	2,3	3,7	3,7
Cinco pontos verbal	3,26	1,01	2,5	4,0	4,0
Cinco pontos não verbal	3,38	1,28	2,0	4,0	4,0
Currículo (Anova – F=1,790, 2 e 246 gl, p=0,168; Levene – F=1,430, 2 e 246 gl, p=0,240)					
Quatro pontos verbal	3,17	1,03	2,3	3,7	3,7
Cinco pontos verbal	3,11	1,03	2,0	3,0	4,0
Cinco pontos não verbal	3,41	1,19	2,0	4,0	4,0
Corpo docente (Anova – F=1,079, 2 e 245 gl, p=0,342; Levene – F=1.956, 2 e 245 gl, p=0,144)					
Quatro pontos verbal	3,25	0,88	2,3	3,7	3,7
Cinco pontos verbal	3,46	0,99	3,0	4,0	4,0
Cinco pontos não verbal	3,42	1,12	2,0	4,0	4,0

Fonte: Dados da pesquisa (2017).

Para complementar essa análise, optamos por proceder ao teste de comparação desses parâmetros, e utilizamos (dentro de cada variável) os testes clássico de análise de variância para médias entre as escalas e o teste de Levene para variâncias. Os resultados, que estão na mesma Tabela 2 sempre ao lado do nome da variável, mostram que, apesar da diferença amostral, não há evidência de diferença estatística nas medidas, por escala (nos dois testes, a hipótese nula é de que não há diferença por tipo de escala).

Como segunda conclusão do estudo, podemos então afirmar que **a verificação de satisfação por escala de Likert alcança resultados semelhantes em termos de medidas de posição e dispersão, independentemente do número de pontos e de a escala ser verbal ou não verbal**. Esse resultado é relevante, em particular, para pesquisas que envolvem esse tipo de escala para teste de associação entre construtos, como é o caso da maioria das pesquisas acadêmicas.

4.3 Análise Das Escalas Diretas De Satisfação

Analisamos nesse item as três últimas escalas, que são aquelas nas quais o respondente é perguntado de forma direta sobre seu nível de satisfação. Como apontam algumas análises da literatura especializada (COSTA, 2011), a escala de Likert, embora útil para pesquisas e testes de relação entre construtos latentes, tem a desvantagem de fazer uma verificação indireta, pois não capta, por exemplo, a satisfação, mas sim a concordância com uma afirmação de satisfação (ou seja, é captada a medida de concordância, e, por essa, a medida do construto). As últimas três escalas aqui analisadas trazem então a verificação direta, com pedido que seja apontado um nível de satisfação de -3 a +3 ou apontando a categoria nominada que melhor se associa ao seu grau de satisfação nos três construtos.

Os resultados para os percentuais e os gráficos estão no Apêndice C. O que mais se destacou foi a proximidade de percentuais nas duas escalas de 7 pontos, em particular que, na escala não verbal, os percentuais somados dos valores +1, +2 e +3 são muito próximos dos percentuais das opções ‘Nem um nem outro’, ‘Um pouco satisfeito’, ‘Bastante satisfeito’ e ‘Extremamente satisfeito’ da escala verbal (em estrutura – 75,6% e 73,7%; em currículo – 78,1% e 76,2%; e em corpo docente – 84,2% e 83,7%). A indicação parece ser de que, na escala não verbal, o ponto central fica na visão do respondente como associado aos valores negativos, e não como ponto neutro, ou seja, as marcações como valores positivos envolvem, na escala verbal, desde o ponto neutro até os valores mais

elevados de satisfação.

Já na escala de 4 pontos, que era verbal e que desde o segundo ponto já se anunciava algum nível de satisfação, o que observamos é que o primeiro ponto (que anunciava ‘nada satisfeito’), corresponde em percentual aos dois primeiros níveis das duas escalas de 7 pontos (que acumulam até em no máximo 10%). Como esperado, a escala com mais pontos tende a ‘dispersar mais’ as âncoras mais restritas de escalas com poucos pontos. Observamos, por outro lado, que a soma dos dois últimos pontos de cada escala tende a se aproximar (os percentuais somados dos dois últimos pontos, respectivamente, para a escalas de sete pontos verbal, não verbal, e de 4 pontos, foram: 50,0%, 50,0% e 47,5% para estrutura; 50,0%, 47,2% e 51,9% para currículo; e 61,0%, 56,1% e 60,7% para corpo docente). Desse modo, se temos a intenção de fazer uma correspondência, seria então a de que, **em escalas de aferição direta com 4 e 7 pontos, os dois últimos pontos comportam valores de percentuais correspondentes nos três tipos de escalas.**

Se a intenção for fazer uma correspondência com a escala binária, não encontramos em qualquer das possibilidades de escalas, percentuais de ‘sim’ correspondentes na escala de satisfação (o que mais se aproxima, mas, com percentuais menores, são justamente os dois últimos pontos somados, conforme observamos pelos resultados da Tabela 1). Ou seja, **se temos a intenção de usar procedimentos que requerem conversão de escalas intervalares diretas de satisfação em escalas binárias (como é o caso da TRI, por exemplo), não há referência segura para tal procedimento**, de modo que não temos como fixar, com maior segurança, um ponto de corte melhor estabelecido.

Assim como procedemos na análise das escalas de Likert, também aqui adotamos o entendimento de que os níveis de satisfação poderiam ser convertidos em quantidades para operacionalização na modalidade de escalas intervalar, para extração de medidas de posição e dispersão. Para fixar o mesmo referencial de amplitude, as escalas foram convertidas todas para o padrão de 1 a 7 (a fórmula de conversão, na escala de bipolar não verbal, que foi de -3 a +3, foi pela fórmula $N = O + 4$; na escala unipolar verbal de 4 pontos, tomados os valores de 1 a 4, usamos a fórmula $N = 1 + 2(O - 1)$; em ambos os casos, N é o novo valor e O é o valor na escala original). Os resultados de médias, desvios padrões e quartis estão indicados na Tabela 3. Considerando somente os resultados descritivos já é possível observar uma variação dos valores das medidas de posição, o que dá sinalização de discrepância nas extrações, com maiores diferenças oriundas sempre da escala de 4 pontos, que são sistematicamente menores. Provavelmente, isso decorre da forma de apresentação da escala, que, nos enunciados, começa a sinalizar satisfação a partir do segundo nível.

Tabela 3: Medidas por escalas de satisfação.

Escalas	Média	Desvio	Quartis		
			Q1	Q2	Q3
Estrutura (Anova – F=15,10, 2 e 239 gl, p<0,001; Levene – F=0,341, 2 e 239 gl, p=0,711)					
Quatro pontos unipolar	3,88	1,56	3,0	3,0	5,0
Sete pontos verbal	4,81	1,54	3,0	5,0	6,0
Sete pontos não verbal	5,03	1,42	5,0	5,5	6,0
Currículo (Anova – F=15,20, 2 e 238 gl, p<0,000; Levene – F=1,244, 2 e 238 gl, p=0,290)					
Quatro pontos unipolar	4,11	1,38	3,0	5,0	5,0
Sete pontos verbal	4,99	1,45	4,0	5,0	6,0
Sete pontos não verbal	5,26	1,27	5,0	5,5	6,0
Corpo docente (Anova – F=16,79, 2 e 238 gl, p<0,001; Levene – F=1,825, 2 e 238 gl, p=0,163)					
Quatro pontos unipolar	4,34	1,49	3,0	5,0	5,0
Sete pontos verbal	5,31	1,34	5,0	6,0	6,0
Sete pontos não verbal	5,46	1,12	5,0	6,0	6,0

Fonte: Dados da pesquisa (2017).

Para termos um entendimento melhor sobre essa diferença, procedemos aos testes de análise de média e variância, novamente com os testes ANOVA e Levene. Pelos resultados, que estão sumarizados na Tabela 3, ao lado do nome das variáveis, observamos pela ANOVA que, nos três construtos, a escala gera diferenças de média, ou seja, podemos rejeitar a afirmação de que as médias dos três tipos de escalas são idênticas. Por outro lado, o mesmo

ocorre com as variâncias dado que, pelo teste de Levene, não podemos rejeitar a igualdade de variâncias de cada extração. **A sinalização é de que a escala tem um efeito sobre a média, mas não sobre a dispersão relativa a essa média.**

Como detalhamento dessa análise, procedemos em complemento da ANOVA o teste post-hoc Tukey HSD, que analisa para os diferentes pares de grupos tendo a hipótese nula de que as médias são idênticas. Pelos resultados do R, que optamos por não apresentar para evitar excesso de dados, tivemos a evidência de que as diferenças existem, nos três construtos, sempre oriundas da diferença da média da escala de 4 pontos em relação às outras duas (nas três comparações, as médias das escalas de 7 pontos são sempre idênticas). Ou seja, a alternativa com 4 pontos em escalas de verificação direta gerou resultados discrepantes em relação às duas alternativas com 7 pontos, porém essas últimas duas escalas, independente do formato de apresentação (daqui verbal e não verbal), têm resultados semelhantes para medidas de posição e dispersão.

Como conclusão dessa terceira análise, temos que, **diferente da estratégia utilizada para mensuração com escala de Likert, o número de pontos parece influenciar as medidas de posição (quando comparamos a escala de 4 pontos com as duas de 7 pontos); no entanto, as escalas com iguais números de pontos, independente de apresentação verbal ou não verbal, geram resultados convergentes em média e em dispersão.**

4.4. Discussão Complementar

Pelos resultados, e assim como no estudo de Dolnicar e Grün (2013), a escala binária completa apresenta resultados de níveis mais altos de satisfação em relação à escala binária afirmativa, porém apenas nos aspectos de estrutura e corpo docente (não ocorreu diferença estatística em currículo). Cabe realçar que uma ‘não resposta’ não significa exatamente uma negação (no caso, insatisfação) e pode simplesmente ser uma ausência de posição. Desse modo, a aferição por ‘escala de afirmação’, por deixar opção para o não posicionamento do respondente, permite que ele não tenha o esforço de refletir e formar julgamento. Por esse entendimento, se a intenção de uma pesquisa é captar a real posição (e não vacilações ou indisposição de resposta), certamente na aferição é melhor não dar a opção de ‘não resposta’. No Quadro 2 temos uma síntese da comparação entre as escalas testadas e considerações sobre as opções.

Quadro 2: Síntese das comparações entre as escalas.

Escalas comparadas	Testes utilizados	Considerações
Escala binária por afirmação vs Escala binária completa	Teste de Proporções	Resultados diferem; ao que parece, a escala binária completa requer mais atenção na resposta e gera resultados mais seguros
Escala Likert 4 pontos verbal vs Escala Likert 5 pontos verbal vs Escala Likert 5 pontos não verbal	ANOVA e Levene	Resultados semelhantes em termos de medida de posição e dispersão nas três escalas Likert.
Unipolar de 4 pontos vs Bipolar de 7 pontos verbal vs Bipolar de 7 pontos não verbal	ANOVA, Levene e post-hoc Tukey HSD	Resultados semelhantes em termos de medida de posição e dispersão apenas nas escalas de sete pontos (verbal e não verbal).

Fonte: Elaborado pelos autores (2018).

Levando em conta as frequências das escalas de Likert, destacamos que, nas escalas de de 4 e 5 pontos, o ponto extremo de concordância recebe mais respostas quando não é verbalizado (com concordância total). Já nas escalas de 7 pontos realçamos o fato de que o ponto declarado neutro (nem concordo nem discordo) absorve mais respostas, podendo ser relativa à opção dada de se abster de resposta, evitando maior esforço para responder o questionário, conforme aponta a literatura (COSTA, 2011; DOLNICAR; GRUN, 2013). Na comparação de percentuais com a escala de 4 pontos sem ponto neutro, a sinalização é de que o neutro corresponde a uma resposta negativa que o respondente opta por não emitir. Se considerarmos ainda que numa escala de concordância a neutralidade pode não ter sentido (segundo Costa (2011), neutro não é grau de concordância, mas sim ‘não resposta’), e que é possível estabelecer uma associação da opção pela neutralidade com níveis negativos de concordância, entendemos que a recomendação de Costa (2011) de que devemos evitar ponto neutro fica melhor enfatizada, tanto pela lógica que o autor propõe quanto pelo resultado empírico observado.

Ainda em relação às escalas de Likert, considerando as medidas descritivas de posição e dispersão na padronização realizada, é destacado o fato de não haver diferença significativa, o que converge com o resultado de Felix (2011). Desse modo, se a operacionalização estatística for baseada nessas medidas, para os pesquisadores não tem diferença entre adotar uma alternativa ou outra. Isso, em verdade, é resultado que encontra fundamentos em outras

referências, e está associado ao fato de a escala Likert apresentar bom desempenho e boa adaptabilidade. Por exemplo, no estudo de Silva Júnior e Costa (2014) o desempenho, considerando as medidas agregadas, da Escala de Likert foi similar ao da escala *phrase completion*, apontada por Hodge e Gillespie (2003, 2007) como mais intuitiva e fácil de responder.

Comparando-as com as escalas binárias, percebemos boa correspondência entre os pontos positivos nas escalas Likert, principalmente naquela de 4 pontos, e a resposta afirmativa na escala binária completa, o que sinaliza a possibilidade de conversão sem maiores problemas quanto à correspondência de respostas entre elas, caso se tenha a opção de uso de técnicas que requeiram esse caminho de análise. Esse resultado converge ainda com o estudo de Dolnicar e Grün (2013).

5. Considerações Finais

Nesse estudo buscamos analisar empiricamente os usos de 8 tipos de escalas de verificação, com diferenças em composição semântica e de pontos. No geral, os resultados permitiram verificar nuances, variações e vantagens ou desvantagens dos diferentes tipos de escalas. Os resultados, expostos no item 4 desse artigo, permitiram o alcance pleno do objetivo definido.

Nosso estudo traz uma contribuição na esfera metodológica das pesquisas em ciências sociais e comportamentais. Nesses campos, os estudos no tema particular de teoria da mensuração buscam debater e investigar formas de minimizar os erros oriundos de uso de métricas e na recomendação de práticas mais apropriadas para as diferentes pesquisas quantitativas realizadas. A finalidade desse tipo de análise, que foi, em boa medida, uma finalidade posta para esse estudo, é compreender como o uso da métrica pode impactar na captação das respostas e, dessa forma, apontar formas mais eficientes de medição.

Os resultados aqui encontrados poderão, portanto, servir de referência para pesquisadores nesses campos (ciências sociais e comportamentais), assim como para docentes que venham a desenvolver atividades de formação em metodologia, pesquisa quantitativa ou teoria da mensuração. Além desses, nossa expectativa é que o estudo seja de utilidade para profissionais do campo de pesquisa de mercado e de opinião, que poderão encontrar aqui indicações de efeitos de usos de métricas como as aqui avaliadas.

O estudo apresentou algumas limitações, com destaque especial em relação à restrição do construto utilizado (analisamos somente satisfação). Caberia analisar resultados em outros construtos, nos quais pode haver variação de efeito da escala. Além disso, a amostra com alunos de educação superior foi adequada a esse construto, porém o uso de respondentes com formação educacional mais avançada e como idades mais concentradas na faixa jovem, sempre pode apresentar variação de resultados, quando é feita comparação com respondentes de menor nível educacional. Outra limitação da pesquisa consiste no número de respondentes por escala, que variou de 80 a 84 respondentes. Ainda que se considere um número bom para as análises aqui procedidas, recomenda-se que novos estudos busquem amostras maiores para verificar possíveis efeitos decorrentes desse fator. Fica ainda a sugestão de comparação de variações de métricas que captem o possível efeito do nível educacional ou de outras variáveis de classificação sociodemográfica e econômica relevante (como sexo, nível de renda, idade etc.).

Referências

- ALCÂNTARA, V. C.; LUIZ, G. V.; FERREIRA, A. C.; TEODORO, S. A. S. Dimensões e determinantes da satisfação de alunos em uma instituição de ensino superior. **Revista Brasileira e Marketing - REMark**, v. 11, n. 3, p. 193-220, 2012.
- ALEXANDRE, J. W. C.; ANDRADE, D. F. D.; VASCONCELOS, A. P. D.; ARAÚJO, A. M. S. D. Uma proposta de análise de um construto para medição dos fatores críticos da gestão pela qualidade por intermédio da Teoria da Resposta ao Item. **Gestão e Produção**, n. 9, v. 2, p. 129-141, 2002.
- ANDRADE, A. L.; CRUZ, R. M.; PAUL, S.; BITENCOURT, R. F. Construção de escalas de diferencial semântico: medida de avaliação de sons no interior de aeronaves. **Avaliação Psicológica**, n. 8, v. 2, p. 197-208, 2009.
- BARBOZA, S. I. S.; CARVALHO, D. T.; SOARES NETO, J. B.; COSTA, F. J. Variações de Mensuração pela Escala de Verificação: uma análise com escalas de 5, 7 e 11 pontos. **Teoria e Prática em Administração**, n. 3, v. 2, p. 99-120, 2013.
- BURNETT, J. J.; DUNE, P. M. An appraisal of the use of student subjects in marketing research. **Journal of Business Research**, n. 14, v. 4, p. 329-343, 1986.
- CARVALHO, D. L. T.; COSTA, F. J.; SOUZA, J. J. B. Variações de Mensuração e Resultado em Pesquisas com Coleta de Dados por Questionário *On-line* e Impresso. **Revista Brasileira de Pesquisas de Marketing, Opinião**

e **Mídia**, n. 17, p. 66-8, 2015.

CASTRO, A. K. S. S. **Evasão no ensino superior: um estudo no curso de psicologia da UFRGS**. 2012. 118 f. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

COSTA, F. J. **Mensuração e desenvolvimento de escalas: aplicações em administração**. Rio de Janeiro: Ciência Moderna, 2011.

DALMORO, M.; VIEIRA, K. M. Dilemas na construção de escalas tipo Likert: o Número de itens e a disposição influenciam nos Resultados? **Revista Gestão Organizacional**, n. 6, p. 161-174, 2014.

DAWES, J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. **International Journal of Market Research**, n. 50, v. 1, p. 61-77, 2008.

DOLNICAR, S.; GRÜN, B. “Translating” between survey answer formats. **Journal of Business Research**, n. 66, p. 1298-1306, 2013.

FEITOSA, M. C.; SOARES, L. S.; BELEZA, C. M.; SILVA, G. R.; LEITE, I. R. Uso de escalas/testes como instrumentos de coleta de dados em pesquisas quantitativas em enfermagem. **Revistas de Políticas Públicas**, n. 13, v. 2, p. 92-97, 2014.

FELIX, R. The impact of scale width on responses for multi-item, self-report measures. **Journal of Targeting Measurement and Analysis for Marketing**, n. 19, v. 3, p. 153-164, 2011.

FINKELSTEIN, L. Widely-defined measurement: An analysis of challenges. **Measurement**, n. 42, p. 1270-1277, 2009.

GARLAND, R. The mid-point on a rating scale: Is it desirable. **Marketing bulletin**, n. 2, v. 1, p. 66-70, 1991.

HODGE, D. R.; GILLESPIE, D. F. Phrase completion: an alternative to Likert scales. **Social Work Research**, n. 27, v. 1, p. 45-5, 2003.

HODGE, D. R.; GILLESPIE, D. F. Phrase Completion Scales: A Better Measurement Approach than Likert Scales? **Journal of Social Service Research**, n. 33, v. 4, p. 1-12, 2007.

LOPES, I. B.; GIRALDI, J. M. E. Uma análise das escalas de mensuração do construto “imagem de país”. **Revista Administração em Diálogo**, n. 16, v. 2, p. 142-166, 2014.

LUCIAN, R. Repensando o Uso da Escala Likert: Tradição ou Escolha Técnica? **Revista Brasileira de Pesquisas de Marketing, Opinião e Mídia**, n. 18, p. 13-32, 2016.

MARI, L. Epistemology of measurement. **Measurement**, n 34, v 1, p 17-30, 2003.

PETERSON, R. A. On the use of college students in social science research: insights from a second-order meta-analysis. **Journal of consumer research**, n. 28, v. 3, p. 450-461, 2001.

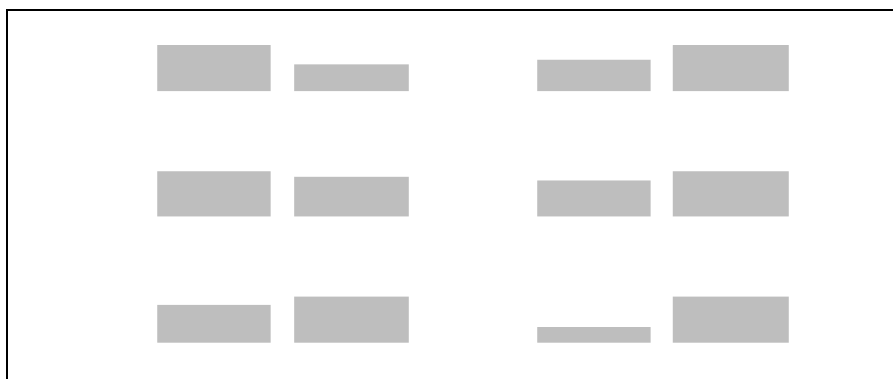
RIFE, S. C.; CATE, K. L.; KOSINSKY, M.; STILLWELL, D. Participant recruitment and data collection through Facebook: the role of personality factors. **International Journal of Social Research Methodology**, n. 19, v. 1, p. 69-83, 2014.

SILVA JÚNIOR, S. D.; COSTA, F. J. Mensuração e Escalas de Verificação: uma Análise Comparativa das Escalas de Likert e *Phrase Completion*. **Revista Brasileira de Pesquisas de Marketing, Opinião e Mídia**, v. 15, p. 1-16, 2014.

UNNITHAN, A. B. Number of Scale Points and Data Characteristics: An Experimental Investigation. **Indian Journal of Science & Technology**, n. 9, v. 1, p. 1-6, 2016.

WARD, P.; CLARK, T.; ZABRISKIE, R.; MORRIS, T. Paper/pencil versus on-line data collection: an exploratory study. **Journal of Leisure Research**, n. 44, v. 4, p. 507-530, 2012.

Apêndice A – Gráficos das escalas binárias

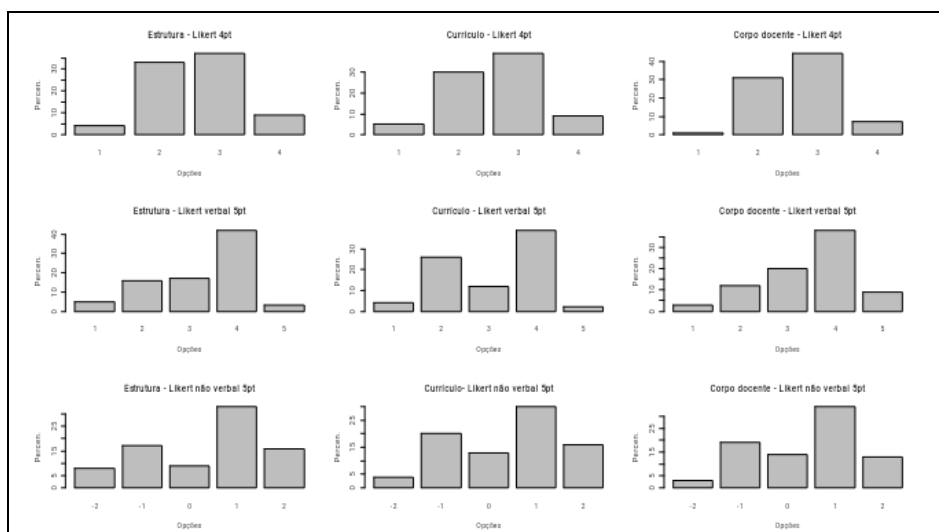


Apêndice B – Tabela de frequência e gráficos das escalas de Likert de 4 e 5 pontos

Escala de cinco pontos não verbal (de -2 a +2)					
Variáveis	-2	-1	0	+1	+2
Estrutura	9,6	20,5	10,8	39,8	19,3
Currículo	4,8	24,1	15,7	36,1	19,3
Corpo docente	3,6	22,9	16,8	41,0	15,7

Escala de cinco pontos verbal					
Variáveis	Discordo totalmente	Discordo	Nem concordo nem discordo	Concordo	Concordo totalmente
Estrutura	6,0	19,3	20,5	50,6	3,6
Currículo	4,8	31,3	14,5	47,0	2,4
Corpo docente	3,7	14,6	24,4	46,3	11,0

Escala de quatro pontos verbal				
Variáveis	Discordo totalmente	Discordo	Concordo	Concordo totalmente
Estrutura	4,8	39,8	44,6	10,8
Currículo	6,0	36,1	47,0	10,9
Corpo docente	1,2	37,3	53,0	8,5



Apêndice C – Tabela de frequência e gráficos das escalas diretas de satisfação

Bipolar de 7 pontos não verbal							
Variáveis	-3	-2	-1	0	+1	+2	+3
Estrutura	1,2	4,9	13,4	4,9	25,6	39,0	11,0
Currículo	1,2	2,4	6,1	12,2	28,1	37,8	12,2
Corpo docente	0,0	0,0	11,0	4,8	23,2	48,8	12,2

Bipolar de 7 pontos verbal							
Variáveis	Extremam. insatisfeito	Bastante insatisfeito	Um pouco insatisfeito	Nem um nem outro	Um pouco satisfeito	Bastante satisfeito	Extremam. satisfeito
Estrutura	2,5	7,5	16,3	7,5	25,0	32,5	8,7
Currículo	0,0	5,0	18,8	7,5	21,3	36,2	11,2
Corpo docente	0,0	5,0	11,3	1,3	26,3	42,4	13,7

Unipolar de 4 pontos verbal				
Variáveis	Nada satisfeito(a)	Um pouco satisfeito(a)	Bastante satisfeito(a)	Extremamente satisfeito(a)
Estrutura	10,0	42,5	42,5	5,0
Currículo	3,8	44,3	44,3	7,6
Corpo docente	5,1	34,2	49,4	11,3

